REVIEW OF THE QUALITY OF A.I.D. EVALUATIONS FY 1987 AND FY 1988

A.I.D. EVALUATION OCCASIONAL PAPER NO. 19
(Document Order No. PN-ABC-321)

by

Paul J. Hopstock
(Development Associates)

Allan C. Kellum
(Development Associates)

Malcolm B. Young, Team Leader
(Development Associates)

U.S. Agency for International Development

May 1989

TABLE OF CONTENTS

LIST OF TABLES

11. Bureau and the Time Required

12. Completeness of Evaluation Report

13. Scope of Work Compliance with A.I.D. Guidelines

14. Information in Evaluation Summaries

15. Completeness of Evaluation Documents by Bureau

16. Completeness of Evaluation Documents by Sector

17. Completeness of Evaluation Documents by Timing of Evaluation

18. Cost of Evaluation by Bureau

19. Cost of Evaluation by Time

20. Cost of Evaluation by Type

21. Team Composition by Bureau

22. Contractor Type by Bureau

23. Congruence of Actionable Evaluation Reports and PES/ES Recommendations

24. Reasons Actions Recommended in Evaluation Report were Excluded from PES/ES Facesheet

25. PES/ES Comments on Quality/Utility of Evaluation Report by Bureau

26. Adequacy of Financial and Program Monitoring of Projects

27. Financial and Program Monitoring by Bureau and Sector

28. Evaluation Reports With Recommendations on Monitoring and Evaluation

29. Percentage of Reports Addressing WID Issues

30. Extent of Participant Training in Projects Evaluated

31. Treatment of Participant Training Topics in Evaluation Reports

32. Methods Used in A.I.D. Evaluations

33. Use of Comparison or Control Groups in A.I.D. Evaluations

FOREWORD

This report is the third review of A.I.D. evaluation studies sponsored by the Bureau for Program and Policy Coordination/ Center for Development Information and Evaluation (PPC/CDIE). The review covers 287 reports submitted by USAID Missions and offices during FY 1987 and FY 1988. Unlike the two previous reviews, which synthesized substantive "lessons learned" from the findings of evaluation reports, this review focuses on the quality of the evaluations as revealed by the characteristics of the studies, including their scope, focus, methods, and techniques.

The review has given PPC/CDIE an opportunity to assess the extent to which the Agency has begun to comply with guidance in the A.I.D. Evaluation Handbook issued in April 1987. Information about the reports submitted during FY 1987 serves as a baseline for tracking the Agency's performance in particular areas that may require special attention and support.

Some of the findings of the review confirm what we already suspected from anecdotal information and from our reading of particular studies. The review gives us a more complete picture of our evaluation practice and, thus, a clearer understanding of its important dimensions. This understanding alerts us to the possibilities for using evaluation more effectively and efficiently and for bringing into our operations the most recent developments in the evaluation field.

For many years, A.I.D. was in the forefront of donor agencies in its support for, and use of, systematic evaluations. More than a dozen countries have come to adopt a version of A.I.D.'s Logical Framework, a technique originally developed by the Agency in the early 1970s to support project planning and evaluation. However, during the last decade, the practice of evaluation in the United States and worldwide has undergone many changes, and the Agency has fallen behind in its efforts to apply and build on new approaches that seem most appropriate and useful for foreign assistance programs.

Other findings of the review alert us to new problems, for example, the declining level of participation on evaluation teams by A.I.D. staff and host country representatives, compared with earlier years. The review also reopens the question of what constitutes a useful, actionable recommendation for Missions and for other entities involved in A.I.D.-supported programs. Finally, A.I.D.'s continuing difficulty in capturing evaluative information on cross-cutting issues takes on new significance in an environment in which accountability for results is increasingly emphasized. For example, the evaluations reviewed still tend to ignore issues related to gender and environmental effects. An exception in this regard is the increased attention devoted to questions about sustainability.

Our readers are welcome to share with PPC/CDIE their comments on the report and their suggestions for future reviews.

Janet Ballantyne
Associate Assistant Administrator
Bureau for Program and Policy Coordination
Center for Development Information and Evaluation
U.S. Agency for International Development
May 1989

SUMMARY

This study presents the results of a review of 287 evaluation reports submitted by U.S. Agency for International Development (A.I.D.) Missions and offices during FY 1987 and FY 1988. It focuses on two main areas:

-- Various measures of compliance with guidance in the A.I.D. Evaluation Handbook (April 1987)

-- Various descriptors of the quality of the evaluations as evidenced in the reports, including their scope, focus, methods, and techniques

Findings

-- Eighty-nine percent of the reports evaluated single A.I.D. projects, 10 percent evaluated more than one project, and 1 percent evaluated nonproject assistance.

-- On certain key measures (completeness of report elements and complexity of evaluation methods), evaluations completed in the Asia and the Near East region and the Latin America and the Caribbean region were rated more positively than evaluations for other regions and bureaus.

-- A.I.D. staff participated as evaluation team members in 29 percent of the evaluations, 53 percent of the evaluations were conducted solely by contracted evaluators, and host country evaluators participated in 27 percent of the evaluations.

-- Sixty-nine percent of the evaluations were interim; that is, they were carried out during project implementation rather than at the end of the project or after project termination.

-- In terms of the primary focus of the evaluation, 64 percent primarily addressed questions about the project's outputs; 28 percent primarily addressed questions about the project's purposes; and 2 percent primarily addressed questions about goals.

-- Almost complete or complete data were available on project outputs in 51 percent of the reports, on project purposes in 19 percent of the reports, and on project goals in 4 percent of the reports. These ratings were generally consistent across sectors and bureaus.

-- Data collection techniques relied heavily on key informant interviews and, to a somewhat lesser extent, on-site visits; little or no use was made of focus group or community interviews, informal or formal surveys, or direct observation. This may reflect the short duration of the evaluations, which averaged about 1 month for fieldwork and preparation of the first draft of the report.

-- Of evaluations using various analytical methods, 11 percent made some use of comparison or control groups; 50 percent analyzed some trend data (over two or more points in time); and 23 percent undertook a detailed cost-effectiveness analysis.

-- Sixty percent of the reports contained information on the project's financial monitoring and 79 percent contained information on program monitoring: on a five-point scale, 66 percent of the projects evaluated

rated high (i.e., at the top two scale points) on the adequacy of financial monitoring, and 54 percent rated high on the adequacy of program monitoring.

-- A total of 59 percent of the reports called for some form of improvement in the project's monitoring, evaluation, or management information systems.

-- Two-thirds of the reports included the required Evaluation Summary or Project Evaluation Summary, with a section listing actions to be taken based on the evaluation.

-- On average, somewhat fewer than half of the recommendations in the evaluation reports were were considered "A.I.D.-actionable; that is, they could be acted on by the sponsors of the evaluations. The remaining recommendations were directed toward those implementing the projects.

-- Forty-two percent of evaluation reports cited prior evaluations. Of those, 42 percent (18 percent overall) noted that recommendations from earlier evaluations had not been implemented.

-- Women-in-development issues were addressed in detail in 9 percent of the evaluations, environmental issues were addressed in detail in 8 percent, and sustainability issues were addressed in detail in 36 percent of the evaluations.

-- Cost data (available for 45 percent of the evaluations) indicate a mean cost per evaluation of $37,130, with 17 percent of the evaluations having costs less than $10,000, and 20 percent having costs more than $60,000.

GLOSSARY

AFR - Bureau for Africa, A.I.D.

A.I.D. - U.S. Agency for International Development

A.I.D./W - U.S. Agency for International Development/ Washington

ANE - Bureau for Asia and Near East, A.I.D.

CDIE - Center for Development Information and Evaluation, A.I.D.

ER(s) - evaluation reports

FY - fiscal year

FVA - Bureau for Food for Peace and Voluntary Assistance, A.I.D.

LAC - Bureau for Latin America and the Caribbean, A.I.D.

M&E - monitoring and evaluation

PES/ES - USAID-Prepared Project Evaluation Summary or Evaluation Summary

PPC - Bureau for Program and Policy Coordination

PRE - Bureau for Private Enterprise, A.I.D.

PVO - private voluntary organization

SOW - scope of work/statement of work

WID - Women in Development

1. INTRODUCTION

1.1 Background and Purpose

The Center for Development Information and Evaluation in the Agency for International Development's Bureau for Program and Policy Coordination (A.I.D./PPC/CDIE) develops and issues

Agency guidance on program and project evaluation, while operational responsibility for the conduct of A.I.D. evaluations is decentralized among the Agency's bureaus and Missions. PPC/CDIE also serves as the repository of A.I.D.'s evaluation-related information and uses its store of data to summarize, synthesize, and disseminate development information of value to managers, planners, and policymakers. CDIE's information base stems largely from A.I.D.'s evaluation reports and their accompanying Project Evaluation Summaries (PESs) or Evaluation Summaries (ESs).

To help accomplish its mission, CDIE periodically under-takes a synthesis and analysis of the Agency's evaluations. This report is a continuation of that effort, which began in 1982. It is based on evaluation data received by CDIE in the past 2 years and focuses on two areas:

-- Measures of compliance with guidance in the 1987 A.I.D. Evaluation Handbook

-- The emphases of the evaluations and the methodologies and techniques they employed

1.2 Methods and Procedures Used

In August 1988 CDIE contracted with Development Associates, Inc. to prepare a written report on the quality and coverage of the evaluation reports submitted by A.I.D. units during FY 1987 and FY 1988, relative to a list of predefined elements. In late August, CDIE furnished Development Associates with a list of evaluation reports to be included in the study.

The process of assembling the needed materials proved to be a difficult one. Many reports are on the CDIE data system without a summary, and some summaries are on the system without any corresponding report; occasionally the same documents have been entered into the system twice, or the report and summary of the same evaluation have been assigned different system identification numbers. Although such anomalies are not unusual in large, complex databases such as CDIE's the lesson to be learned here is simply that the CDIE evaluation database is not yet working perfectly, and users should approach it with that realization.

Simultaneous to assembling the materials for review, Development Associates, in consultation with CDIE, refined the list of elements to be assessed and developed a rating and data entry form for recording the presence, absence, value, extent, or degree of the information of interest. The resulting form (contained in Appendix A) provided for more than 100 discrete entries from each evaluation.

Once the materials were assembled and the data recording form finalized and approved by A.I.D., the raters were given an orientation on the coding procedures. The rating process was divided into two steps. First, an initial rater recorded the factual elements that involved little professional judgment or required little detailed knowledge of A.I.D. Then, a more senior and experienced professional read the evaluation report and the accompanying PES/ES to rate the remaining, more substantive items, such as identifying the principal focus of the evaluation analysis. Two individuals were responsible for completing Step 1, and four individuals participated in Step 2.

During the initial weeks of the rating period the raters met frequently to clarify their interpretations and discuss the treatment of unusual cases to ensure a high degree of reliability in the ratings. In addition, Step 2 raters checked the Step 1 ratings, and a random set of 50 reports was rated independently by all possible pairs involved in completing Step 2. Once a high degree of reliability was established, the remainder of the forms were completed first by one and then by a second rater.

Once ratings were complete, the forms were thoroughly edited, and the data entered into a dBASE III+ file, using a customized data entry screen identical to the rating/data entry form. In addition, dBASE III+ was used to calculate values (e.g., the "life of project") using variables related to the project's start and end dates and a sequence of dBASE codes.

Verification of a significant random sample of the data resulted in the statistical assurance that the data entry process had an accuracy of more than 99 percent. While dBASE III+ was used for the initial data entry, calculations, and refining procedures, SYSTAT was the application software used for the analytical procedures and for calculating composite variables. Lotus 1-2-3 was used for generating graphs and sorting the file to present a list of all the projects and evaluation reports (see Appendix B).

1.3 Contents of This Report

The purpose of this report is to present descriptive findings and selected analyses from the evaluation synthesis. Section 2 provides an overview of the focus and coverage of the evaluations included in this study. Section 3 focuses on the evaluation management process; it addresses the time required to complete the steps in the evaluation process, the completeness of evaluation scopes of work and reports, evaluation costs, the characteristics of implementors of the evaluations, and the evaluations' perceived utility to A.I.D. Section 4 presents information about previous monitoring and evaluations of projects evaluated, and Section 5 provides data on three cross-cutting issues of interest to CDIE (i.e., women in development, the environment, and participant training). The final section presents data on the methods and techniques used in A.I.D. evaluations. The appendixes include the rating form and instructions and a list of the projects covered by the evaluation reports.

2. FOCUS AND COVERAGE OF EVALUATIONS

A total of 287 evaluation reports were examined, of which 255 (89 percent) were evaluations of single projects, 29 (10 percent) evaluated more than one project, and 3 (1 percent) did not evaluate projects, but rather examined other forms of program assistance (e.g., housing guarantees).

The evaluation reports examined can be described in terms of the characteristics of (1) the projects and (2) the evaluation processes and reports.

2.1 Characteristics of Projects Evaluated

Because most of the evaluations were of single projects, a summary of project characteristics can be made. In some cases reports concerned multiple projects or no projects; in these cases the project characteristics are listed as "missing" in tables.

Table 1 shows the number of evaluated projects by bureau. As can be seen, 87 percent of the evaluations were of projects in regional bureaus, although there were also a significant number of projects financed by the Science and Technology Bureau. Table 2 shows the regional locations of the projects evaluated, and the most frequent country locations. Projects in Honduras, Costa Rica, and Egypt were the most frequently evaluated.

Table 1. Bureau of Projects Evaluated

Bureau	No. of Evaluation Reports	Percentage
Latin America and the Caribbean	97	34
Asia and the Near East	79	28
Africa	73	25
Science and Technology	29	10
Food for Peace and Voluntary Assistance	5	2
Private Enterprise	4	1
Total	287	100

Table 2. Regional and Most Frequent Country Locations of Projects Evaluated

Region/Country	No. of Evaluation Reports	Percentage
Africa	79	28
Zaire	9	3
Lesotho	7	2
Asia and the Near East	84	29
Egypt	11	4
Bangladesh	9	3
Indonesia	8	3
Thailand	8	3
India	7	2
Pakistan	7	2
Latin America and the Caribbean	109	38

Honduras	13	5
Costa Rica	11	4
Bolivia	9	3
Peru	8	3
Ecuador	7	2
Multiregion	15	5

Raters categorized projects according to sectors, funding size, and length of project. As indicated in Table 3, health and population projects were evaluated most frequently, followed by rural development and agriculture projects. Analyses relating sector and bureau indicated that Africa had a higher than average number of agricultural projects, Latin America and Caribbean had a higher than average number of private enterprise projects, and the Science and Technology Bureau had a higher than average number of projects involving health and population.

Table 3. Projects Evaluated by Sector

Sector	No. of Evaluation Reports	Percentage
Health and Population	76	26
Rural Development	53	18
Agriculture	50	17
Private Enterprise	40	14
Forestry, Energy, Environment, and Natural Resources	27	9
Education and International Training	21	7
Nutrition	8	3
Urban Development	5	2
Other	7	2
Total	287	100

Note: Percentage totals in this and other tables may not add to 100 because of roundings.

As shown in Table 4, the most frequent functional accounts were Agriculture, Rural Development and Nutrition, and Economic Support Fund. Funding size and length of project (in years) are shown in Tables 5 and 6.

Table 4. Projects Evaluated by Account

Account	No. of Project Evaluations	Percentage
Agriculture, Rural Development, and Nutrition	67	30
Economic Support Fund	40	18
Health	29	13
Selected Development Activities	29	13
Population	24	11
Education and Human Resources	17	8
Sahel Development Program	13	6
International Disaster Relief/Assistance	4	2
Southern Africa Fund	2	1
Child Survival	1	0
Total	226	100

Note: Reports missing this characteristic = 61 (21 percent).

Table 5. Projects Evaluated by Funding Size

Funding Size	No. of Projects	Percentage
Less than $500,000	10	4
$500,000-$1 million	13	6
$1 million-$5 million	50	22
$5 million-$10 million	47	21
$10 million-$50 million	91	40
$50 million-$100 million	11	5
More than $100 million	5	2
Total	227	100

Note: Reports missing this characteristic = 60 (21 percent).

Table 6. Projects Evaluated by Project Length

Project Length (Years)	No. of Evaluation Reports	Percentage
2	16	7
3	20	9
4	33	14
5	47	20
6	30	13
7	25	11
8	27	12
9	12	5
10 or more	21	9
Total	231	100

Note: Reports missing this characteristic = 56 (20 percent).

Finally, evaluation reports were examined to determine whether evaluators had rated the project or its components as "highly successful." (Typically, evaluators are not asked to render an overall summative assessment, so an absence of comment does not imply a negative judgment.) A total of 41 evaluation reports (14 percent) reported a highly successful project, and an additional 32 reports (11 percent) reported a highly successful project component.

Also, an analysis was conducted to determine the types of projects reported to be highly successful or to have highly successful components. Agriculture projects (14 percent) and education and training projects (15 percent) were less likely than other projects (31 percent) to be evaluated as highly successful. Also, the longer the project, the more likely it would be rated highly successful (2-4 years = 16 percent, 5-9 years = 24 percent, 10 or more years = 52 percent). There were no major differences by bureau or project size.

2.2 Characteristics of Evaluation Reports

The evaluation reports were categorized according to whether they were interim evaluations (carried out more than 6 months prior to project completion), final evaluations (carried out in the last 6 months of the project or within 1 year after project completion), ex post evaluations (carried out more than a year following project completion), or other evaluations (not project specific). Most of the reports (69 percent) were interim

evaluations, though there were also a large number of final evaluations (1 percent). Also, there were no major differences in percentages of interim and final evaluations by sponsor or sector.

Evaluation reports were also categorized according to whether they were internal or external evaluations. An evaluation was characterized as internal if it included anyone from A.I.D. or the organization implementing the project on the evaluation team. Using this standard, 31 percent of these evaluations that could be rated were internal. It should be noted that only 7 percent of all evaluations are done using only A.I.D. personnel, so most of the internal evaluations actually involved mixed teams. The proportion of internal evaluations was particularly large in Africa (51 percent) and Asia and Near East (41 percent). Only 13 percent of evaluations sponsored by other bureaus were done internally.

Scopes of work for evaluations and evaluation reports were also examined to determine the extent to which inputs, outputs, purposes, goals, and assumptions were addressed. In each case, the extent of emphasis in the scopes of work or evaluation report was rated on a four-point scale: 0 = not at all; 1 = addressed minimally; 2 = addressed in detail; and 3 = primary focus. Table 7 shows the distributions for these variables.

Table 7. Emphases of Scopes of Work and Evaluation Reports (percentages)

	No. of Evaluation Reports	Not at All (0)	In Minimally (1)	Detail (2)	Primary Focus (3)	Total
Scopes of Work						
Inputs	179	20	50	27	3	100
Outputs	179	1	2	36	61	100
Purposes	179	2	17	51	31	100
Goals	179	27	53	18	3	100
Assumptions	173	21	47	31	1	100
Evaluation Reports						
Inputs	287	4	50	43	3	100

Outputs	287	0	4	32	64	100
Purposes	287	2	18	53	28	100
Goals	282	22	57	18	2	100
Assumptions	259	10	49	40	1	100

As can be seen from Table 7, inputs and assumptions were somewhat more likely to be addressed in evaluation reports than in scopes of work for evaluations. In general, however, evaluation reports seemed to reflect the emphases of the relevant scopes of work.

The emphases of evaluation reports did not differ by sponsor, sector, or timing of the evaluation. For every major category of reports, outputs were most frequently emphasized, followed by purposes.

3. EVALUATION MANAGEMENT PROCESS

3.1 Sequence and Timing of Evaluation Steps

Each of the evaluations included in this report was logged into the PPC/CDIE data system during FY 1987 or FY 1988. However, there was considerable variation in the speed with which each evaluation progressed from one step to the next in A.I.D.'s evaluation process.

3.1.1 Overview of the System

Figure 1 provides an overview of the A.I.D. evaluation process, beginning with the preparation of the statement of work (SOW) and ending with the completion of the actions recommended in the evaluation report. As shown, there are seven steps in the process, and this study obtained information on the time required to complete five of the steps. Since the material available did not include calendar dates for the preparation of the SOW, nor, for the most part, the actual start of the evaluation, it is not possible to estimate the overall start-to-finish calendar time required for the process. However, assuming the time between submission of a draft evaluation report and completion of the report's final version was about 1.5 months, {1} it can be estimated that a typical evaluation required a bit less than 3 months between the time the evaluation team actually began work until a final report was submitted. About 10 months was needed before the evaluation summary was signed and copies

INSERT FIGURE 1

of the report were made available to the rest of the Agency, and the greatest time needed to complete an evaluation's actionable recommendations was 15 months after date of signature.{2}

==========

(1)This assumption is based on Development Associates' experience in conducting over 100 A.I.D. evaluations in the past 6 years.

(2)In obtaining these results, two evaluations that extended for 2 years and a third for 3 years were considered anomalies and excluded from the time and cost calculations. Similarly, three evaluations published in 1979, 1983, and 1984 were excluded from calculations of the time elapsed before the reports were entered in the CDIE database.

3.1.2 Time Required for Specific Steps in the Evaluation

The time required to complete each step in the evaluation management process was analyzed in terms of (1) the date of the evaluation report, (2) the type of evaluation (internal or external), (3) timing (interim or final), and (4) sponsoring bureau. The results of these analyses are provided below.

Date of Evaluation and the Time Required. The evaluation reports were divided into two categories based on the calendar year on the cover page of the evaluation report. Those dated 1986 or before (n=90) constituted one group, and those dated 1987 or after (n=161) made up the other. In addition, there was a group of undated reports (n=36), which were excluded from the analyses.

Table 8 shows the time requirements for the various evaluation steps for the two time periods. There were significant improvements in the rate of completion of each evaluation step following the completion of the evaluation report. While the time needed for the evaluation itself remained virtually the same, the other time intervals decreased dramatically. The largest improvement was the reduction of 5.3 months in the second step, the time from evaluation report publication to completion of the PES/ES.

Type of Evaluation and the Time Required. As indicated in Table 9, there was little variation between internal and external evaluations in the time required for the evaluation steps. Internal evaluations required somewhat less time for entry into the CDIE file.

Table 8. Date of Evaluation Report and the Time Required

Step	1986 or before (months)	1987 or after (months)
Duration of Evaluation	1.1 (n=68)	1.2 (n=105)
From Final Report to Director's Signature on PES/ES (reports with summaries)	8.7 (n=56)	3.4 (n=79)
From Director's Signature to Entry Onto PPC/CDIE File (reports with summaries)	4.1 (n=53)	2.4 (n=76)
From Final Report to Entry Onto PPC/CDIE File (all reports)	9.7 (n=89)	5.2 (n=161)
From Director's Signature to Most Distant Recommended Action (reports with summaries)	9.0 (n=42)	6.0 (n=63)

Table 9. Type of Evaluation and the Time Required

Step	Internal (months)	External (months)
Duration of Evaluation	1.0 (n=59)	1.2 (n=134)
From Final Report to Director's Signature on PES/ES (reports with summaries)	6.1 (n=40)	5.4 (n=95)
From Director's Signature to Entry Into PPC/CDIE File (reports with summaries)	2.0 (n=46)	3.6 (n=99)
From Final Report to Entry Into PPC/CDIE File (all reports)	5.7 (n=73)	7.3 (n=174)
From Director's Signature to Most	7.5	6.5

Distant Recommended Action (reports with summaries)	(n=41)	(n=76)

Timing of the Evaluation and the Time Required. Table 10 shows the mean time for each of the two main timing categories -- interim and final. Final evaluation reports took 1.5 months longer from date of report to entry into CDIE file than those for interim evaluations. Likewise, it took an average of 2.6 months longer to complete all the recommended actions listed in the ESs for final evaluations than those for interim evaluations.

Table 10. Timing of the Evaluation and the Time Required

Step	Interim (months)	Final (months)
Duration of Evaluation	1.1 (n=139)	1.1 (n=55)
From Final Report to Director's Signature on PES/ES (reports with summaries)	5.7 (n=90)	5.5 (n=43)
From Director's Signature to Entry Onto PPC/CDIE File (reports with summaries)	2.6 (n=96)	4.4 (n=48)
From Final Report to Entry Onto PPC/CDIE File (all reports)	6.4 (n=173)	7.9 (n=73)
From Director's Signature to Most Distant Recommended Action (reports with summaries)	6.1 (n=84)	8.7 (n=31)

Bureau and the Time Required. As Table 11 shows, the relative performance of the bureaus was examined for each step in the evaluation management process. The lengthier duration of evaluations in the "other" category, unusually high at 4.5 months (n=2), was largely due to the Bureau for Private Enterprise. Perhaps of most interest in the table is that evaluations completed in Latin America and the Caribbean took the longest to enter the CDIE system.

Table 11. Bureau and the Time Required

Step	AFR (months)	ANE (months)	LAC (months)	Other (months)
Duration of Evaluation	1.0 (n=42)	1.1 (n=60)	1.1 (n=71)	1.5 (n=22)
From Final Report to Director's Signature on PES/ES (reports with summaries)	7.6 (n=25)	3.7 (n=45)	6.4 (n=59)	3.5 (n=6)
From Director's Signature to Signature on PES/ES (reports with summaries)	2.8 (n=29)	2.7 (n=46)	3.9 (n=67)	1.0 (n=5)
From Final Report to Entry Onto PPC/CDIE File (all reports)	6.2 (n=63)	5.9 (n=73)	8.4 (n=82)	5.9 (n=32)
From Director's Signature to Most Distant Recommended Action (reports with summaries)	6.0 (n=24)	8.6 (n=39)	5.9 (n=51)	4.5 (n=4)

The average length of time required to complete all the actions listed in the ESs -- i.e., including the most "distant" actions -- was greatest in the case of evaluations in Asia and the Near East.

3.2 Completeness of Various Elements

The evaluation reports, the ESs, and the evaluation SOWs were each examined for completeness of key elements, and composite ratings were developed for each report. Each composite represents the total number of elements present out of 16 possible for each ES, and 8 apiece for evaluation reports and SOWs. The ESs were more frequently complete (38 percent had composite ratings of 16). By contrast, only 10 percent of the evaluation reports and 5 percent of the SOWs were complete (composite ratings of 8).

The completeness of the 287 evaluation reports averaged 5.3 on the composite rating scale of 0 to 8 based on the eight features listed in Table 12. As the table shows, 90 percent of the evaluations contained recommendations, the feature most frequently included. The project's Logical Framework appeared or was discussed in fewer than one-third of the evaluation

reports.

Table 12. Completeness of Evaluation Report
(percentages)

Feature	Included in Evaluation Report: No	Yes
Executive Summary	21	79
Table of Contents	13	87
Evaluation SOW	46	54
Methodology	24	76
Conclusions	29	71
Recommendations	10	90
Lessons Learned	61	39
Logical Framework	68	32

Note: Number of evaluation reports reviewed was 287; composite rating = 5.3

SOW compliance with A.I.D. guidelines was judged against the checklist of the eight features listed in Table 13. On a scale of 0 to 8, 8 indicating full compliance, the mean rating was 4.6. The two features that appeared most frequently (83 percent) were the statement of purpose and the list of study questions. Only 10 percent of evaluation SOWs contained the required funding section. However, the SOWs were often edited or incomplete versions of the original evaluation SOWs -- a factor that should be considered in interpreting what otherwise would appear to be an extremely low level of compliance.

Table 13. Scope of Work Compliance with A.I.D. Guidelines
(percentages)

Feature	Addressed in SOW: No	Yes
Activity To Be Evaluated	35	65
Purpose of Evaluation	17	83
Background Information	54	46
SOW Study Questions	17	83

Methods and Procedures	37	63
Team Composition	43	57
Reports Required	44	56
Funding	90	10

Note: Evaluation reports with SOWs = 156; composite rating = 4.6

ESs were completed for 117 of the evaluation reports. In addition, PESs were completed for 68 reports, and no summaries were completed for 102 reports. The completeness rating of the 117 ESs had a mean of 14.7 on a 16-point scale based on the 16 features listed in Table 14. Thus, the ESs were found to be highly complete. Only one feature, lessons learned, appeared with a frequency less than 85 percent, and five features appeared with individual frequencies of 95 percent or greater. The low frequency (65 percent) with which lessons learned were included in the ESs may be largely attributed to the absence of separately labeled "lessons learned" sections in many of the evaluation reports themselves.

In order to examine the factors related to completeness of evaluation reports, a series of linear multiple regressions were performed using the composite of report completeness. The factors included in prediction equations were sponsor, sector, type of evaluation, timing of evaluation, date of evaluation, length of evaluation, evaluation cost, and focus of evaluation report. The results indicated that sponsor, sector, and timing of evaluation were related to report completeness, but that the other factors were not. Thus, the evaluation report completeness composite as well as the other two composites were examined by bureau, sector, and timing. Composite ratings by bureau are presented in Table 15.

Table 14. Information in Evaluation Summaries (percentages)

Features Included	No	Yes
Reporting A.I.D. Unit	9	91
Was Evaluation Scheduled in Current FY Annual Evaluation Plan?	3	97
Evaluation Timing	8	92
Activity or Activities Evaluated	0	100

Action Decisions Approved by Mission or A.I.D./W Office Director	0	100
Date of Mission or A.I.D./W Office Review of Evaluation	11	89
Approvals of ES and Action Decisions	0	100
Evaluation Abstract	1	99
Evaluation Costs	6	94
Purpose of Activity Evaluated	8	92
Purpose of Evaluation and Methodology Used	10	90
Findings and Conclusions	7	93
Recommendations	11	89
Lessons Learned	35	65
Attachments	14	86
Comments by Mission, A.I.D./W and Borrower/Grantee on Evaluation Report	10	90

Note: No. of reports with ES = 117; composite rating = 14.7

Table 15. Completeness of Evaluation Documents by Bureau

	Mean Composite Ratings		
Bureau	Evaluation Report	Evaluation SOW	Evaluation Summary
AFR	4.7 (n=73)	4.3 (n=30)	13.7 (n=9)
ANE	5.6 (n=79)	5.0 (n=52)	14.5 (n=56)
LAC	5.7 (n=97)	4.5 (n=48)	15.2 (n=46)
Other	4.6 (n=38)	4.3 (n=16)	13.8 (n=6)

Total	5.3 (n=287)	4.6 (n=156)	14.7 (n=117)

Table 16 presents the composite ratings by sector. Evaluations in the private enterprise, urban development, and energy/environment/national resources sectors had evaluation reports that were more complete than the average.

The completeness of evaluation documents was analyzed according to timing of evaluation. The results in Table 17 show that final evaluations tended to have higher completeness indicators than interim evaluations.

3.3 Evaluation Costs by Bureau, Timing, and Type

Evaluation costs were reported on 130 of the reports studied. As stated earlier, three cases were excluded from these cost analyses because of the unusually high costs associated with very lengthy evaluations. Also, for the purpose of data analysis, the evaluation costs denominated in host country currencies were converted to U.S. dollars by using the exchange rate for the approximate date of evaluation completion. The mean evaluation cost was $37,131, with 17 percent of the evaluation having costs less than $10,000, and 20 percent having costs more than $60,000. Table 18 presents cost data by bureau.

Table 16. Completeness of Evaluation Documents by Sector

	Mean Composite Ratings		
Sector	Evaluation Report	Evaluation SOW	Evaluation Summary
Agriculture	5.1 (n=50)	4.7 (n=25)	14.8 (n=18)
Rural and Institutional Development	5.0 (n=53)	4.0 (n=27)	14.6 (n=18)
Health and Population	5.2 (n=76)	4.0 (n=43)	14.7 (n=31)
Nutrition and PL 480 Title II	4.1 (n=8)	3.3 (n=3)	14.7 (n=4)

Education and International Training	4.8 (n=21)	5.0 (n=7)	14.7 (n=10)
Private Enterprise	6.1 (n=40)	5.5 (n=26)	14.7 (n=21)
Forest, Energy, Environment and National Resources	5.7 (n=27)	5.2 (n=17)	14.1 (n=9)
Urban Development	6.8 (n=5)	4.0 (n=2)	14.7 (n=3)
Other	6.3 (n=7)	6.0 (n=6)	15.7 (n=3)
Total	5.3 (n=287)	4.6 (n=156)	14.7 (n=117)

Table 17. Completeness of Evaluation Documents by Timing of Evaluation

	Mean Composite Ratings		
Timing	Evaluation Report	Evaluation SOW	Evaluation Summary
Interim	5.2 (n=198)	4.5 (n=111)	14.6 (n=81)
Final	5.5 (n=84)	5.0 (n=44)	15.0 (n=34)

Table 18. Cost of Evaluation by Bureau

	No. of	Cost in U.S. Dollars		
Bureau	Evaluation Reports	Mean	Minimum	Maximum
AFR	13	31,798	2,000	90,000
ANE	50	39,174	1,250	109,400
LAC	55	36,654	1,400	185,904
Other	9	40,900	8,601	107,568

Total	127	37,131

Cost was also examined based on the evaluation timing. As Table 19 shows, the average cost of the interim evaluations exceeds that for final evaluations by approximately 14 percent.

Table 19. Cost of Evaluation by Time

Timing of Evaluation	No. of Evaluation Reports	Cost in U.S. Dollars Mean	Minimum	Maximum
Interim	89	39,033	1,250	109,400
Final	37	34,381	1,400	185,904

The costs for internal, as opposed to external, evaluations were also compared. As Table 20 shows, external evaluations cost only slightly more than internal evaluations.

Table 20. Cost of Evaluation by Type

Timing of Evaluation	No. of Evaluation Reports	Cost in U.S. Dollars Mean	Minimum	Maximum
Internal	35	36,974	2,000	109,400
External	92	37,631	1,250	185,904

Finally, the relationship between evaluation cost and length of evaluation was examined. Not surprisingly, significant correlation (r=.36) between them was found.

3.4 Characteristics of Evaluation Teams and Contractors

The implementers of the evaluations can be described in terms of (1) the composition of the evaluation teams and (2) the type of contractor.

The evaluations, based on the composition of the evaluation teams, were divided into six categories. Table 21 shows the six categories and the results of the analysis of team composition and bureau sponsorship. The results suggest that the Africa evaluations were most likely to use A.I.D. personnel and least likely to use contractors.

There were no significant team composition differences for interim versus final evaluations.

The evaluations were also categorized by the type of contractor -- defined as the responsible organization for conducting the evaluation, generally the organization supplying the team leader. The contractor types consist of three main categories: (1) U.S. contractors, (2) U.S. personal services, and (3) non-U.S. contractors. If the evaluation was led by A.I.D. personnel, "not applicable" was coded.

Table 21. Team Composition by Bureau
(percentages)

Team Composition	Bureau AFR	ANE	LAC	Other	Total
A.I.D. Only	15	5	3	3	7
Contractor Only	33	43	67	71	53
Host Country Only	3	4	12	3	6
A.I.D. and Contractor	16	13	5	13	11
A.I.D. and Host Country	1	5	0	0	2
Contractor and Host Country	15	15	5	5	10
A.I.D. and Contractor and Host Country	15	14	2	3	9
Indeterminant	1	1	4	3	2
Total	100 (n=73)	100 (n=79)	100 (n=97)	100 (n=38)	100 (n=287)

As shown in Table 22, the centrally funded projects were found to use U.S. contractors most frequently, and the Africa evaluations used U.S. contractors least frequently.

Table 22. Contractor Type by Bureau
(percentages)

Type of Contractor	Bureau AFR	ANE	LAC	Other
U.S. Contractor	36	55	67	84
U.S. Personal Services	20	18	10	5
Non-U.S. Contractors	8	14	15	3
Not Applicable	28	13	4	5
Don't Know	5	1	3	3
Total	100 (n=74)	100 (n=78)	100 (n=97)	100 (n=38)

3.5 Perceived Utility of Results

A.I.D.'s perception of the utility of an evaluation report can be judged by two criteria: (1) It can be inferred on the basis of A.I.D.'s reaction to an evaluation report's recommendations, and (2) it can be perceived more directly from the PES/ES's comments about the report's quality/utility. The following two subsections analyze the evaluation reports according to these criteria.

3.5.1 A.I.D.-Actionable Recommendations

The congruence of A.I.D.-actionable recommendations in the evaluation report with those on the PES/ES facesheet was rated to infer the utility of the evaluation report. Two other variables were important in defining this variable: (1) a simple count of the number of actions listed on the PES/ES facesheet and (2) an assessment of the percentage of evaluation report recommendations that were actionable by A.I.D. and thus eligible for inclusion on the PES/ES.

Of the 185 evaluation reports with summaries, 10 percent were judged as having no A.I.D.-actionable recommendations. At the other extreme, all of the evaluation report recommendations were actionable for 14 percent of the reports. On average, somewhat fewer than half of the recommendations were actionable.

No significant variations in the number of actionable recommendations were found by bureau, timing, evaluation date, or type.

Also, the simple count of the number of A.I.D. action decisions appearing on the PES/ES facesheet showed an overall mean value of 6.5 (n=185) with a maximum of 51. Not surprisingly, the mean was higher (7.7, n=124) for interim evaluations than for final ones (4.2, n=58). For internal evaluations the mean was 5.9 (n=123) compared with 7.3 (n=59) for external ones; before 1987 the mean was 7.5 (n=71) and after that date it was 5.9 (n=90).

As noted earlier, the congruence of A.I.D.-actionable recommendations in the evaluation report with those on the ES facesheet was a means of inferring the utility of the report's recommendations. This presumed that the greater the congruence, the greater was A.I.D.'s agreement with the results. The measure of congruence was based on the percentage of actionable recommendations that appeared in the PES/ES facesheet. Thus, if six actionable recommendations were in the evaluation report and four of them appeared in the ES, the percentage was judged to be 75. Percentages were then categorized on the congruence scale that appears in Table 23. As the table shows, the congruence of actionable items between the evaluation report and PES/ES was judged to be "total" or "almost total" for 43 percent of the evaluations.

Table 23. Congruence of Actionable Evaluation Report and PES/ES Recommendations

Congruence	No. of Evaluation Reports	Percentage
None (0%)	28	15
Minimal (1-25%)	16	9
Some (26-50%)	24	13
A Lot (51-75%)	38	21
Almost Total (76-99%)	44	24
Total (100%)	35	19
Total	185	100

Note: Reports missing this characteristic = 102 (36 percent)

The six points on the congruence scale were used to calculate mean values by sponsor, timing, date, type, and sector. The Africa and central bureaus had the highest mean values. Interim evaluations placed higher on the scale than final ones,

evaluations dated 1986 or before had slightly higher congruence than those after 1986, and internal evaluations had higher means than external evaluations. A comparison of the various sectors based on the mean value of their evaluations on the congruence scale yielded no significant differences.

Since the congruence scale is largely a measure of the degree of exclusion of items from the PES/ES that had been deemed actionable by the evaluation team, the reasons for their exclusion were also of interest. Table 24 lists these reasons and the percentages of cases in which they applied.

3.5.2 Comments on Quality/Utility

Of the 185 PES/ESs examined, 111 contained comments on the quality or utility of the evaluation report. These comments fell into three groups: (1) entirely positive; (2) entirely negative; and (3) mixed, containing both positive and negative. Comments (or the lack of a comment) on each of the 185 PES/ESs

Table 24. Reasons Actions Recommended in ERa Were Excluded From PES/ES Facesheet

Reason	No. of Evaluation Reports	Percentage of Casesb
ER recommendations are more specific/ detailed than those of ES	62	41
PES/ES recommendations are more specific/ detailed than those of ER	30	20
Mission/Office said recommendations are impractical or not feasible	19	13
ER recommendations are moot because project ended	19	13
Recommended action already underway/ implemented	7	5
Basis for recommendation(s) questioned/ disputed	5	3
Mission opted for course of action that obviated ER recommendation	5	3
PES/ES actionable items are consistent with		

ER text but not specifically cited as ER recommendation	4	3
Adoption of some ER recommendations eliminated need for others	2	1
No reasons specified/discernible	23	15
Total	176	

aER = evaluation report.
bMore than one reason could be cited. Thus, the number of responses (n=176) is greater than the number of cases with responses (n=150), and the total percentage is greater than 100.

were categorized as follows: 41 percent (n=76) entirely positive, 15 percent (n=28) mixed, 4 percent (n=7) entirely negative, and 40 percent (n=74) with no comment. Analyses by date and by timing showed no significant differences.

Table 25. PES/ES Comments on Quality/Utility of Evaluation Report by Bureau

Bureau	No. of Reports	Type of Comments (percentages)				
		Negative	Mixed	Positive	None	Total
AFR	42	2	2	10	86	100
ANE	58	7	16	52	26	100
LAC	78	3	22	49	27	100
Other	7	0	14	57	29	100
Total	185	3	13	42	42	100

Note: Reports missing this characteristic = 102 (36 percent)

Analyses of the quality/utility comments by bureau revealed significant differences in the percentage of PES/ESs with comments. As shown in Table 25, 14 percent of the Africa PES/ESs contained comments, compared with 73 percent for the

other bureaus.

4. MONITORING AND EVALUATION

Evaluation reports were examined to determine the adequacy of monitoring systems and the presence and outcomes of prior evaluations of the projects. The results on these topics are described in this section.

4.1 Adequacy of Monitoring

Reviewers were asked to rate, on a five-point scale of 0 (wholly inadequate) to 4 (wholly adequate), the adequacy of financial and program monitoring of projects based on comments in the evaluation reports (see Table 26).

Table 26. Adequacy of Financial and Program Monitoring of Projects

	Financial Monitoring		Program Monitoring	
	No. of Reports	Percentage	No. of Reports	Percentage
0 (Wholly inadequate)	2	1	2	1
1	20	7	28	10
2	36	13	74	26
3	84	29	96	33
4 (Wholly adequate)	29	10	28	10
5 (Information unavailable)	116	40	59	21
Total	287	100	287	100

Perhaps the most striking finding from this table is the large number of evaluation reports that did not include any evaluative comments on financial monitoring (40 percent) and program monitoring (21 percent). Financial monitoring (mean = 2.69) was rated somewhat higher than program monitoring (mean = 2.52). In both cases, 3 was the most frequent rating.

Table 27 shows the percentage of missing responses (i.e., data unavailable) and mean ratings by sponsor and sector. Projects in Asia and Near East received the lowest ratings on both financial and program monitoring. Agricultural and energy/environmental projects got the lowest ratings on

financial monitoring, and education/training and energy/ environmental projects got the lowest ratings on program monitoring.

A total of 59 percent of the reports contained recommendations on monitoring and evaluation, management information systems, or information planning (see Table 28). The most frequent recommendation was to upgrade existing information systems using present resources.

4.2 Prior Evaluations of Projects

Forty-two percent of the evaluation reports cited previous evaluations. Final evaluations were cited more often than interim evaluations (58 percent as opposed to 36 percent). Previous evaluations of education and training, rural development, and health and population projects were most likely to be cited. There were no major differences on this item by sponsor.

Table 27. Financial and Program Monitoring by Bureau and Sector

	No. of Reports	Financial Monitoring Data Missing (%)	Financial Monitoring Mean Rating[a]	Program Monitoring Data Missing (%)	Program Monitoring Mean Rating[a]
Bureau					
AFR	73	32	2.72	19	2.66
ANE	79	52	2.50	23	2.34
LAC	97	30	2.72	15	2.55
Other	38	61	2.93	32	2.65
Sector					
Agriculture	50	36	2.31	22	2.62
Rural Development	53	30	2.81	21	2.62
Health and Population	76	47	2.82	25	2.54
Nutrition	8	62	3.00	12	2.57
Education and Training	21	52	2.50	19	2.18
Private Enterprise	40	38	3.12	15	2.68
Energy and Environment	27	41	2.36	19	2.27
Urban Development	5	60	2.50	40	2.67

Other	7	14	2.33	0	2.14

aScale = 0-4

Table 28. Evaluation Reports With Recommendations on Monitoring and Evaluation

Recommendation	No. of Reports	Percentage
Upgrade information systems without additional inputs	56	20
Create new information systems	32	11
Improve coordination/communication between project and A.I.D.	26	9
Add new project inputs (staff, materials) to improve information systems	21	7
Change timing/frequency of evaluations	17	6
Provide outside technical assistance or training to improve information systems	14	5
Other	2	1

Raters also indicated whether reports cited any unimplemented actions suggested by previous evaluations. Eighteen percent of all reports cited unimplemented actions from prior evaluations, a figure that represents 42 percent of the evaluation reports that cited previous evaluations. The percentage of cited evaluations with unimplemented actions was higher for Latin America and Caribbean (53 percent) and Africa (48 percent) than for other bureaus (31 percent).

5. CROSS-CUTTING ISSUES

Three major cross-cutting issues were examined as part of the assessment of A.I.D. evaluation reports: women in development (WID), the environment, and participant training.

5.1 Women in Development

Raters indicated the extent to which WID issues were addressed in evaluation reports using a three-point scale: 0«=«not addressed, 1 = addressed minimally, and 2 = addressed in detail. The overall results showed that WID issues were not addressed in 67 percent of reports, were addressed minimally in 24 percent of reports, and were addressed in detail in only 9 percent of reports.

There were significant differences in the treatment of WID issues by sponsor and sector. Table 29 shows the percentage of reports addressing WID issues by subgroups. The evaluations in Asia and Near East and central bureaus were more likely to address WID issues than were evaluations in Africa and Latin America and Caribbean. WID issues were also particularly likely to be addressed in evaluations of projects in the rural development and education and training sectors.

Table 29. Percentage of Reports Addressing WID Issues

Subgroup	No. of Reports	Not at All (%)	Minimally (%)	In Detail (%)	Total (%)
Bureau					
AFR	73	75	19	5	100
ANE	79	57	28	15	100
LAC	97	71	22	7	100
Other	38	61	32	8	100
Sector					
Agriculture	50	80	16	4	100
Rural Development	53	55	30	15	100
Health and Population	76	66	25	9	100
Nutrition	8	100	0	0	100
Education and Training	21	57	29	14	100
Private Enterprise	40	72	18	10	100
Energy and Environment	27	63	33	4	100
Urban Development	5	80	20	0	100
Other	7	43	43	14	100

5.2 Environment

Raters indicated the extent to which environmental issues were addressed in evaluation reports on the same three-point

scale as was used for WID issues. The overall results showed that environmental issues were not addressed in 75 percent of reports, were minimally addressed in 17 percent of reports, and were addressed in detail in 8 percent of reports. As might be expected, environmental issues were particularly likely to be addressed in reports related to energy, environment, and natural resources (addressed = 78 percent), but less likely to be addressed in reports on other sectors (addressed = 20 percent). There was also a difference by type of evaluation, with internal evaluations (33 percent) more likely to address environmental issues than external evaluations (22 percent).

5.3 Participant Training

The third cross-cutting issue examined in the evaluation reports was participant training. Raters first indicated whether the evaluation report mentioned participant training as a project component, and if so, to what extent (0 = no mention of participant training, 1 = minor component of project, 2 = major component of project, and 3 = entire project was participant training).

As shown in Table 30, 60 percent of evaluation reports did not mention participant training, and only 3 percent of reports concerned projects that consisted entirely of participant training.

Table 30. Extent of Participant Training in Projects Evaluated

Extent	No. of reports	Percentage
No evidence of participant training	173	60
A minor project component	70	24
A major project component	36	13
Entire project was participant training	8	3
Total	287	100

Data on participant training were examined by bureau and sector. The results indicated that projects evaluated in Latin America and Caribbean and the Food for Peace and Voluntary Assistance Bureau were less likely than average to include participant training, while projects evaluated in the Science and Technology and the Private Enterprise bureaus were more likely than average to include participant training as a major

or sole component. Projects in the following sectors were also more likely than average to include participant training: agriculture, health and population, and education and training.

Evaluation reports that addressed participant training were also checked on whether they addressed (1) training management and operations (selection, processing, assignment, support, and follow-up); (2) number of people trained or who completed training; (3) appropriateness of post-training employment/activities; (4) short-term or micro effects of training (e.g., on workplace, colleagues); and (5) long-term or macro effects of training (e.g., on institutional or public policies, economy, targeted beneficiaries). The treatments of these topics were rated on a scale of 0 = not addressed, 1 = addressed minimally, and 2 = addressed in detail. The overall results on these items are shown in Table 31.

Table 31. Treatment of Participant Training Topics in Evaluation Reportsa (percentages)

Topic	Not Addressed (0)	Addressed Minimally (1)	Addressed in Detail (2)	Total
Management and Operations	32	47	20	100
Numbers Trained	22	34	44	100
Post-Training Employment	55	29	16	100
Short-Term Effects	75	21	4	100
Long-Term Effects	74	14	12	100

aNumber of evaluation reports = 114.

Evaluation reports were most likely to discuss the number of participant trainees and least likely to discuss short-term and long-term effects of training. There were no major differences on these variables based on sponsor or timing of evaluation. That is, final evaluations were no more likely than interim evaluations to address the effects of training; the issue was ignored by three-quarters of the evaluations of projects with training components.

6. METHODS AND TECHNIQUES USED IN A.I.D. EVALUATIONS

A series of ratings were made concerning the methods and techniques used in A.I.D. evaluations and the manner in which results were presented. This section describes (1) the specific methods used in evaluations; (2) the availability of data to evaluators to assess aspects of the project; (3) the treatment in the reports of sustainability and unexpected positive and ega negative impacts; and (4) the presentation in A.I.D. evaluations of conclusions and recommendations.

6.1 Methods Used

The evaluation reports reflected a wide variety of approaches to and techniques of data collection. Seven specific techniques (key informant interviews, focus group interviews, community interviews, direct observation, informal surveys, formal surveys, and site visits) were assessed (see Appendix A for detailed definitions of these techniques). For each technique, raters gave a score of 0 to 3 based on the following scale: 0 = not used, 1 = limited use, 2 = extensive use, and 3 = extensive and exemplary use (see Table 32). As can be seen, key informant interviews and site visits were most frequently used. Focus group interviews and community interviews were little used.

Raters also examined reports to determine whether comparison or control groups were used, and rated them on the following scale: 0 = none reported, 1 = unplanned and limited, 2 = unplanned but extensive, 3 = planned but limited, and 4 = planned and extensive (see Table 33). As can be seen, use of comparison groups in evaluations is relatively rare.

Reports were also examined to see whether trend data were used in the analyses of outputs, purposes, or goals. Use of trend data was rated as follows: 0 = none reported; 1 = yes, two points in time (e.g., pre-post); and 2 = yes, three or more points in time. The results are presented in Table 34. This table shows that trend data were used in half of the evaluations.

Table 32. Methods Used in A.I.D. Evaluations
(percentages)

Method	Not Used	Limited Use	Extensive Use	Extensive and Exemplary Use	Total
Key Informant Interviews	11	44	43	2	100

Focus Group Interviews	99	1	0	0	100
Community Interviews	96	4	1	0	100
Direct Observation	73	18	9	0	100
Informal Survey	80	11	8	1	100
Formal Survey	90	3	6	1	100
Site Visits	31	37	31	1	100

Note: n = 284, Missing = 3 (1 percent)

Table 33. Use of Comparison or Control Groups in A.I.D. Evaluations

Use of Control Groups	No. of Reports	Percentage
None reported	254	89
Unplanned and limited	11	4
Unplanned but extensive	2	1
Planned but limited	14	5
Planned and extensive	6	2
Total	287	100

Table 34. Use of Trend Data in A.I.D. Evaluations

Use of Trend Data	No. of Reports	Percentage
None reported	143	50
Yes, two points in time	40	14
Yes, three or more points	104	36
Total	287	100

Finally, raters examined reports to determine whether cost effectiveness analyses were presented. Reports were categorized according to whether the issue was: 0 = not addressed, 1 = addressed minimally, or 2 = addressed in detail. These results are presented in Table 35. They indicate that cost effectiveness was addressed, at least minimally, in approximately 60 percent of reports.

Table 35. Treatment of Cost-Effectiveness in A.I.D. Evaluations

Treatment	No. of Reports	Percentage
None addressed	114	40
Addressed minimally	108	38
Addressed in detail	65	23
Total	287	100

In order to create a measure of the overall methodological complexity of the evaluations, a composite was developed of the 10 items presented in this section. Because formal surveys and use of comparison groups were considered to be particularly complex and difficult, they were double weighted in the composite. The scores on the composite ranged from 0 to 17 and were well distributed, as shown in Table 36.

The factors associated with methodological complexity were examined through the use of a series of multiple linear regression analyses. The variables included in the prediction equations were project sponsor, type of evaluation, sector, primary focus of report, evaluation cost, evaluation duration, evaluation timing, and date of report.

The results of the regressions suggested that bureau and type of evaluation were related to methodological complexity. Interestingly, neither length of evaluation nor evaluation cost was significantly related to methodological complexity. The mean scores for subgroups of evaluation reports based on sponsor and type of evaluation are shown in Table 37.

Table 36. Methodological Complexity of A.I.D. Evaluations (composite)

Score	No. of Reports	Percentage
0	4	1
1	13	5
2	27	10
3	45	16
4	41	14
5	30	11

6	37	13
7	20	7
8	20	7
9	16	6
10	13	5
11	9	3
2-17	9	3
Total	284	100

Table 37. Methodological Complexity by Bureau and Type of Evaluation (composite)

Bureau and Type of Evaluation	Mean Score	No. of Reports
Bureau		
AFR	4.93	73
ANE	5.99	79
LAC	5.62	95
Other	4.58	38
Type of Evaluation		
Internal	4.77	86
External	5.70	195

6.2 Data Availability

The availability of data for evaluation reports to assess outputs, purposes, goals, and assumptions of projects was evaluated. As shown in Table 38, data were most available on outputs and least available on goals.

Table 38. Data Availability Concerning Outputs, Purposes, Goals, and Assumptions

	Data Availability (percentages)						
No. of Reports	None (0)	Minimal (1)	(2)	Some (3)	A lot (4)	Almost Complete (5)	Comp. Total

Outputs	285	0	3	14	32	42	9	100
Purposes	285	1	18	32	31	18	1	100
Goals	274	16	46	24	10	4	0	100
Assumptions	234	4	22	33	36	3	0	100

Table 39 shows the mean ratings for each of the four data availability items based on timing of the evaluation and sector. Slightly more data are available for final evaluations than for interim evaluations. There is generally a high degree of consistency among sectors, as well as across project sponsors and types of evaluation (internal versus external).

6.3 Treatment of Special Issues

Raters examined each report to determine the extent to which it addressed the issues of sustainability, unexpected negative impacts, and unexpected positive impacts. These were rated on a scale of 0 = not addressed, 1 = addressed minimally, and 2 = addressed in detail (see Table 40). Sustainability was a frequently addressed issue, but unexpected positive and negative impacts were infrequently addressed.

Table 39. Data Availability by Timing of Evaluation and Sector

Timing and Sector	No. of Reports	Outputs	Purposes	Goals	Assumptions
Timing					
Interim	162-198	3.33	2.46	1.33	2.12
Final	69-83	3.59	2.63	1.58	2.23
Sector					
Agriculture	43-50	3.24	2.36	1.35	2.23
Rural Devel.	42-53	3.41	2.57	1.55	2.43
Health & Pop.	62-76	3.30	2.43	1.24	1.84
Nutrition	6-8	3.12	2.88	2.14	2.18
Educ. & Train.	14-21	3.29	2.24	1.24	1.93

Private Ent.	33-39	3.67	2.80	1.60	2.18
Energy/Env.	24-26	3.54	2.50	1.67	2.21
Urban Devel.	4-5	3.80	3.60	1.25	2.25
Other	6-7	3.71	2.14	0.57	2.33

a0 = none, 5 = complete

Note: The range of the number of reports reflects some variation in the inclusion of data on all four data items.

Table 40. Treatment of Sustainability and Unexpected Positive and Negative Impactsa (percentages)

	Not Addressed (0)	Addressed Minimally (1)	Addressed in Detail (2)	Total
Sustainability	25	39	36	100
Unexpected Positive Impacts	84	14	2	100
Unexpected Negative Impacts	84	13	3	100

aNumber of Reports = 287

Sustainability was more frequently addressed in reports sponsored by the Latin America and Caribbean Bureau (79 percent) and the Asia and Near East Bureau (78 percent) than in reports sponsored by other bureaus (69 percent). Sector and timing of the report had little bearing on treatment of sustainability while timing of the evaluation was a more important factor in addressing unexpected positive and negative impacts. Final evaluations were more likely than interim evaluations to address unexpected positive impacts (21 percent versus 14 percent) and unexpected negative impacts (23 percent versus 14 percent).

6.4 Presentation of Conclusions and Recommendations

Two items in the review of evaluation reports related to conclusions and recommendations. First, raters were to judge the extent to which findings, conclusions, and recommendations reflected analysis of empirical data. Second, they were to judge the extent to which the evaluation reports appropriately distinguished between conclusions and recommendations. Both items employed a five-point scale ranging from not at all (0) to completely (4).

The overall results on the first item are presented in Table 41. Evaluation reports were generally given high ratings on this item, with 79 percent of reports receiving a rating of 3 or 4.

Table 41. Use of Empirical Data to Generate Findings, Conclusions, and Recommendations

Rating	No. of Reports	Percentage
0 (Not at all)	3	1
1	11	4
2	46	16
3	167	58
4 (Completely)	60	21
Total	287	100

There were minor differences on this item based on sponsor of the evaluation. Evaluation reports from the Asia and Near East (mean = 3.05) and Latin America and Caribbean (mean = 3.00) bureaus were given higher ratings than those from other bureaus (mean = 2.81).

Evaluation reports were also highly rated on appropriate distinctions between conclusions and recommendations (see Table 42). Seventy-four percent of reports received a rating of 3 or 4.

Table 42. Percentage of Evaluation Reports That Appropriately Distinguished Between Conclusions and Recommendations

Rating	No. of Reports	Percentage

0 (Not at all)	8	3
1	28	10
2	38	13
3	96	33
4 (Completely)	117	41
Total	287	100

There were no major differences on this item based on sponsor, sector, or timing of evaluation.